(12) UK Patent Application (19) GB (11) 2 310 718 (13) A

(43) Date of A Publication 03.09.1997

(21) Application No 9703774.1

(22) Date of Filing 24.02.1997

(30) Priority Data
(31) 9604314 (32) 29.02.1996 (33) GB

(71) Applicant(s)
British Textile Technology Group

(Incorporated in the United Kingdom)

Shirley House, Wilmslow Road, Didsbury, MANCHESTER, M20 2RB, United Kingdom

(72) Inventor(s)
Jonathan Mark Appleby
Paul Frederick Hamlyn
Gordon Nelson

(74) Agent and/or Address for Service
McNeight & Lawrence
Regent House, Heaton Lane, STOCKPORT, Cheshire, SK4 1BS, United Kingdom

(51) INT CL[6]
C12Q 1/68

(52) UK CL (Edition O)
G1B BAC B108 B200 B220 B303
U1S S1068 S1274 S1334

(56) Documents Cited
C.R.Acad.Sci.Paris,Sciences de la vie/Life Sciences 1994,317 , 685-692 Mol.Biol.Evol. 1994,11(5),769-777 Mol.Gen.Genet. 1993,239,450-462 Annu.Rev.Plant Physiol.Pant Mol.Biol, 1989,40,415-439 Aust.J.Bot. 1983,31,395-410

(58) Field of Search
UK CL (Edition O) C3H HB4A , G1B BAC
INT CL[6] C12Q 1/68
ONLINE: WPI,CLAIMS,DIALOG/BIOTECH

(54) Determination of plant phylogeny involving introns from the RuBisCo small subunit encoding gene

(57) A process for the determination of the phylogeny of plant genetic material comprises the characterisation of at least one intron from the ribulose 1,5-bisphosphate carboxylase (RuBisCo) encoding gene therefrom, and comparison of said intron(s) with at least one previously-characterised intron from a RuBisCo small subunit encoding gene. The comparison may involve introns having substantially identical flanking sequences, the sequences of the introns, or the length of one or more of the introns (optionally including intervening sequences). The genetic material may be genomic DNA. The process may be used to determine the phylogeny of processed plant genetic material, or the genus, species or variety of plant genetic material.

A preferred process comprises measurement of the length of the introns in a double-stranded nucleic acid by:

(i) separation of the two strands of the nucleic acid, thereby to yield single-stranded molecules; performing, at least once, the following:

(ii) annealing, to the single-stranded molecules (the 'template strands'), oligonucleotide primers specific to conserved sequences adjacent to the intron(s), whose length is to be measured, of the RuBisCo small subunit encoding gene;

(iii) extending the bound oligonucleotide primers to yield complementary copies of the template strands;

(iv) separation of the resultant double-stranded products to yield single-stranded molecules;

wherein the phylogeny of the plant genetic material is determined by measurement of the lengths of the extended oligonucleotide primers, and comparison of the lengths of said extended primers with the length of extended oligonucleotide primers previously measured from at least one intron of a RuBisCo small subunit encoding gene. Two oligonucleotide primers may be used, one specific to a conserved sequence 5' to the first intron, and the other specific to a conserved sequence 3' to the last intron, of a RuBisCo small subunit encoding gene.

GB 2 310 718 A

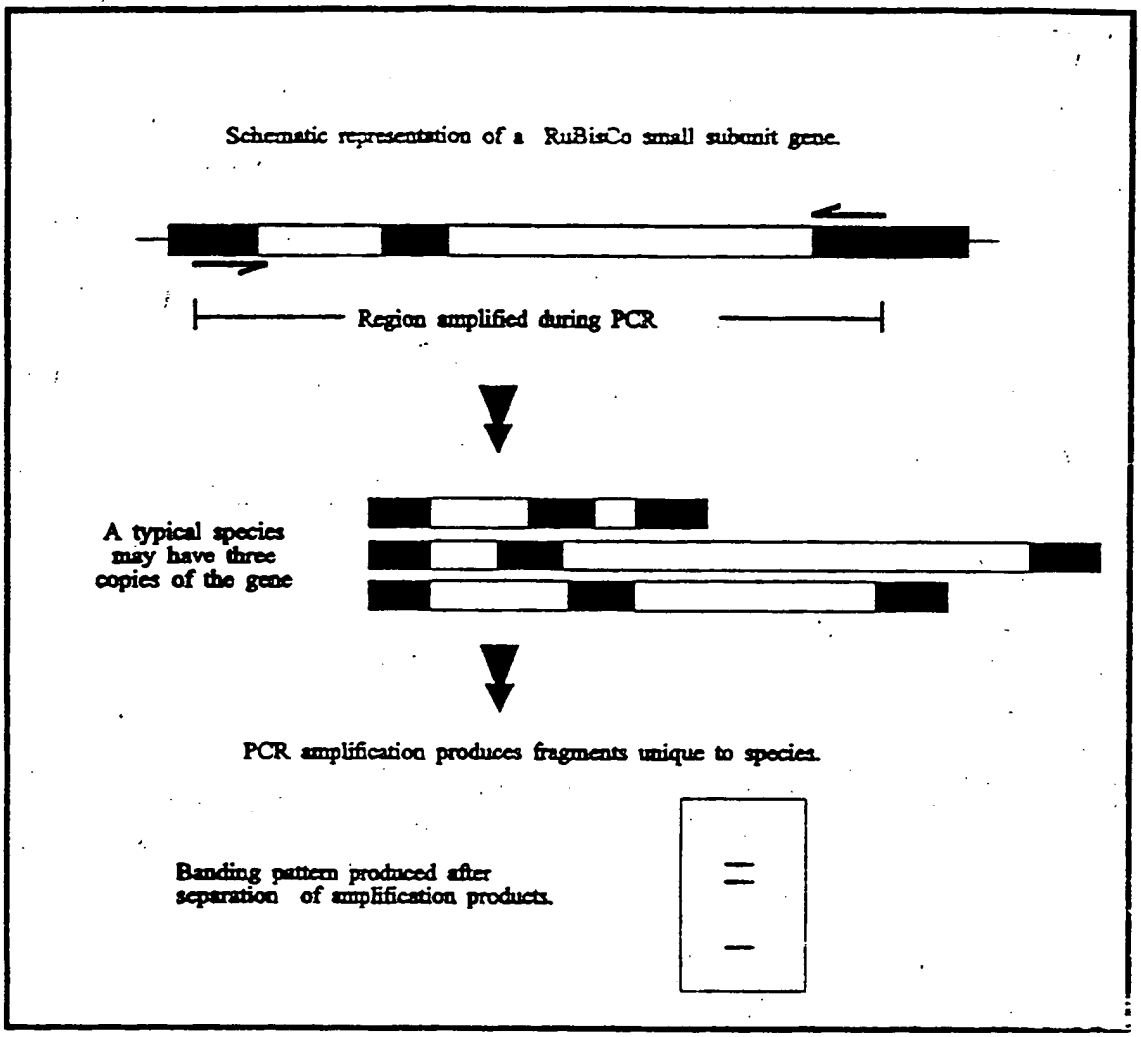

Schematic representation of a RuBisCo small subunit gene.
Region amplified during PCR
A typical species may have three copies of the gene
PCR amplification produces fragments unique to species.
Banding pattern produced after separation of amplification products.

FIGURE 1

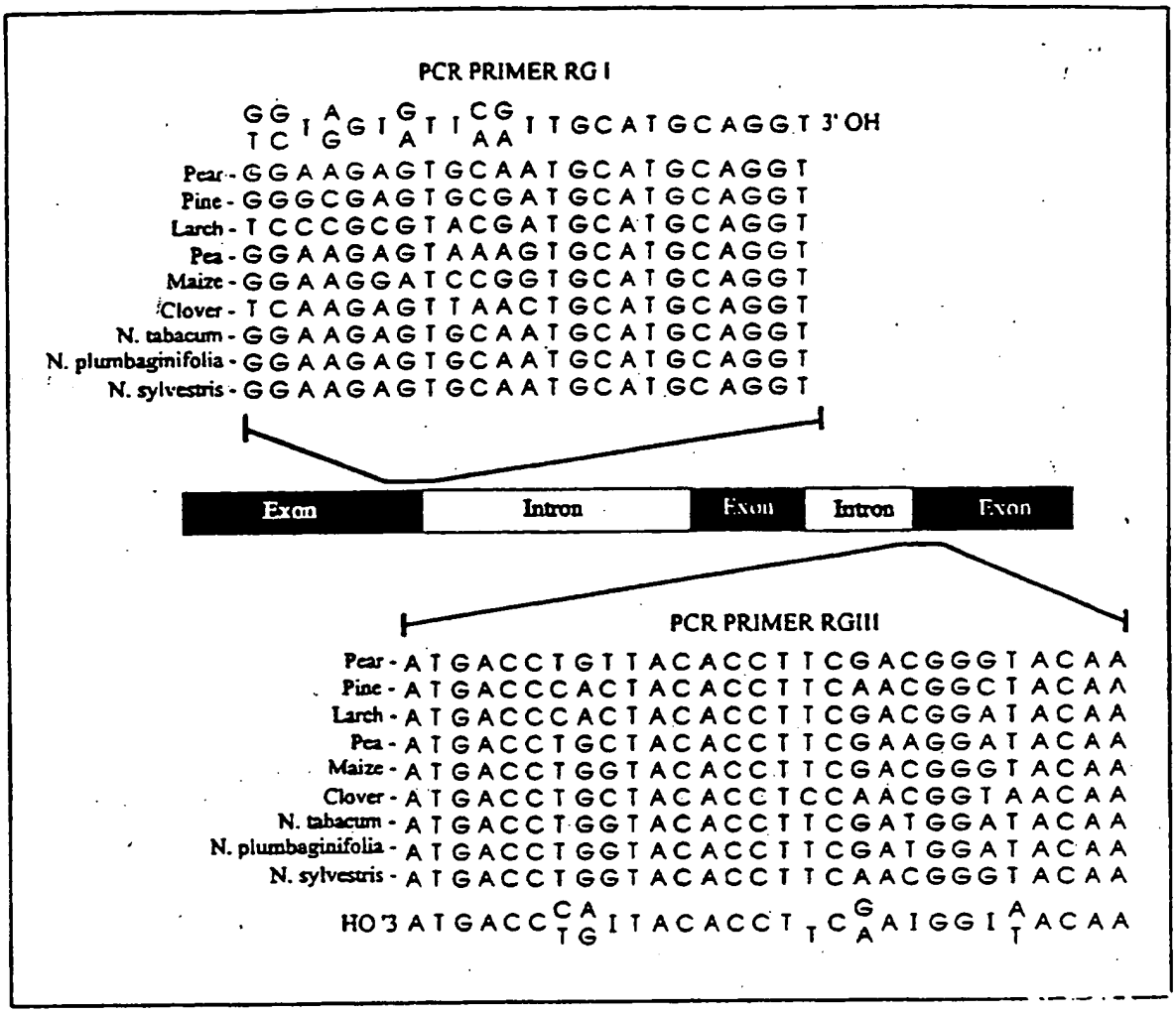

PCR PRIMER RG I
GG/TC I A/G G I G/A T I CG/AA I T GCA T GCAGG T 3' OH
Pear - GGAAGAGTGCAATGCATGCAGGT
Pine - GGGCGAGTGCGATGCATGCAGGT
Larch - TCCCGCGTACGATGCATGCAGGT
Pea - GGAAGAGTAAAGTGCATGCAGGT
Maize - GGAAGGATCCGGTGCATGCAGGT
Clover - TCAAGAGTTAACTGCATGCAGGT
N. tabacum - GGAAGAGTGCAATGCATGCAGGT
N. plumbaginifolia - GGAAGAGTGCAATGCATGCAGGT
N. sylvestris - GGAAGAGTGCAATGCATGCAGGT
Exon
Intron
Exon
Intron
Exon
PCR PRIMER RGIII
Pear - ATGACCTGTTACACCTTCGACGGGTACAA
Pine - ATGACCCACTACACCTTCAACGGCTACAA
Larch - ATGACCCACTACACCTTCGACGGATACAA
Pea - ATGACCTGCTACACCTTCGAAGGATACAA
Maize - ATGACCTGGTACACCTTCGACGGGTACAA
Clover - ATGACCTGCTACACCTCCAACGGTAACAA
N. tabacum - ATGACCTGGTACACCTTCGATGGATACAA
N. plumbaginifolia - ATGACCTGGTACACCTTCGATGGATACAA
N. sylvestris - ATGACCTGGTACACCTTCAACGGGTACAA
HO'3 ATGACC C/T A/G I TACACCT T C G/A AIGGI A/T ACAA

FIGURE 2

Figure 3

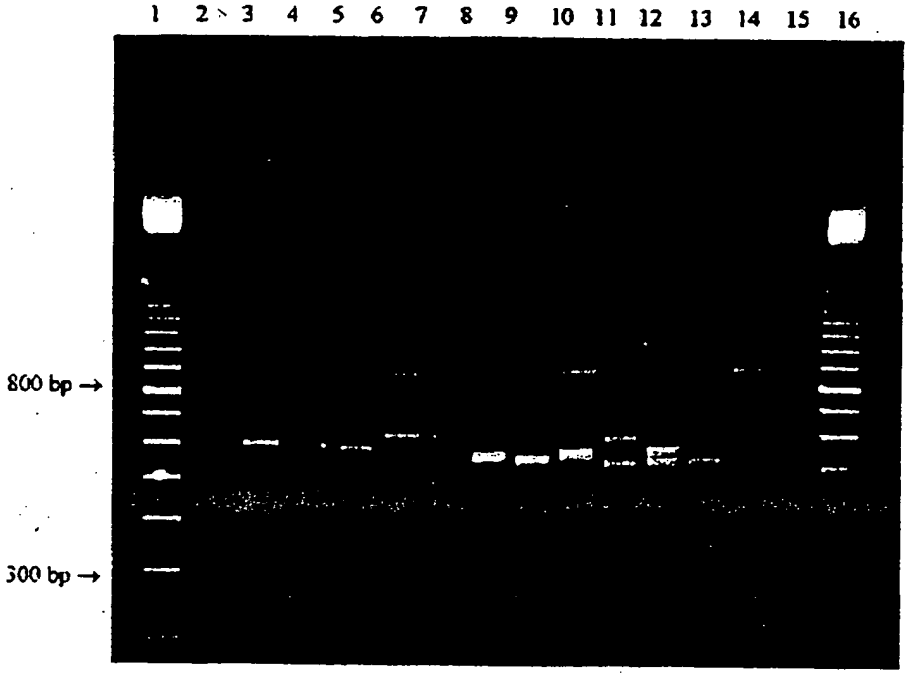


Figure 4

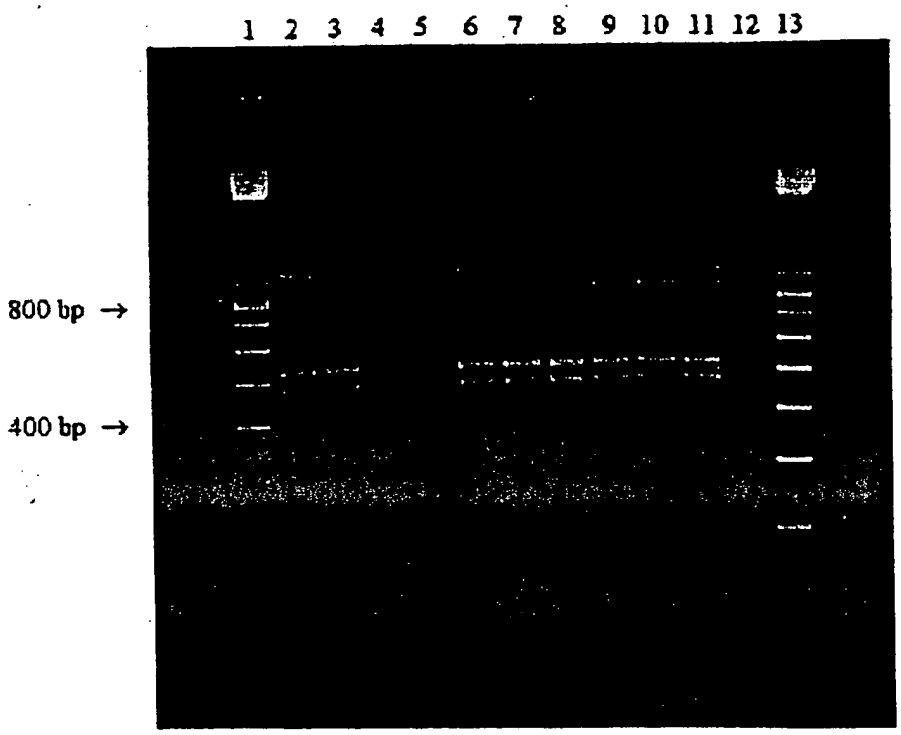


Figure 5

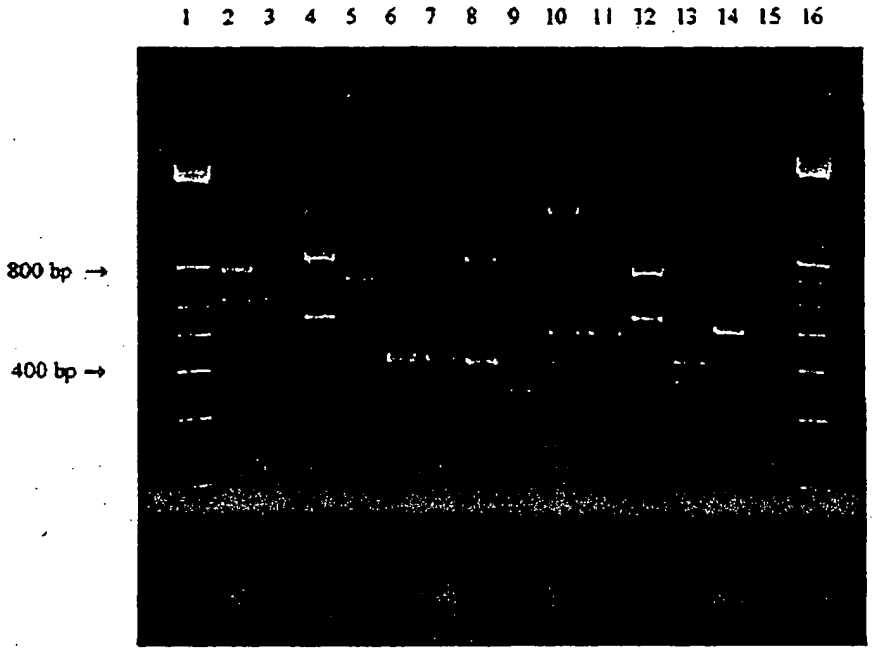
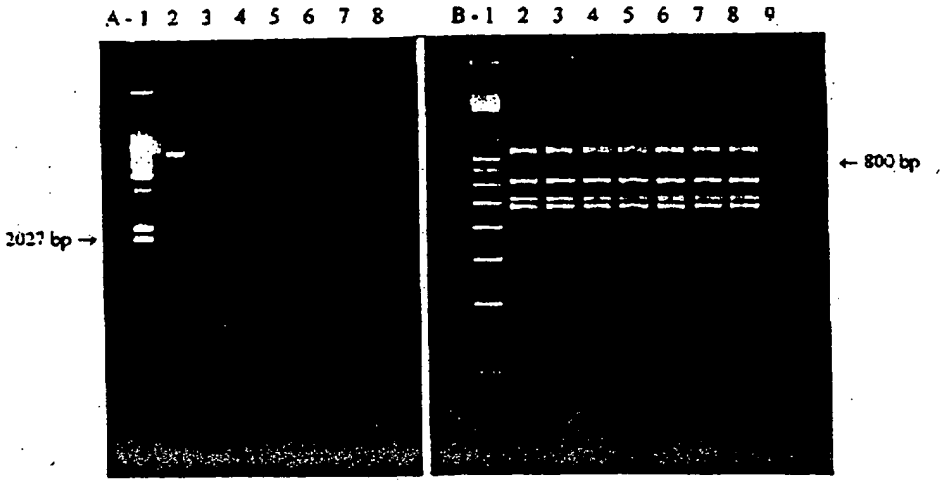


Figure 6



2310718

Plant Phylogenisation

The present invention concerns the phylogenisation of plant genetic material, in particular but not limited to the speciation of plant genetic material.

The identification of the phylogeny, in particular the species, of plant material is very important, e.g. for ensuring the identity and quality of material received by a manufacturer. For example, a cigarette manufacturer may wish to ensure the species of *Nicotiana* used in cigarettes.

Many methods of genetic analysis for tissue identification are currently available, for example genetic fingerprinting (Jeffreys, A.J. *et al.*, 1985, Nature 316: 76-79), RAPD (Random Amplification of Polymorphic DNA) (Williams, J.G.K. *et al.*, 1990, Nucleic Acids Res., 18: 6531-6535; Welsh, J. and McClelland, M., 1990, Nucleic Acids Res., 18: 7213-7218), FINS (Forensically Informative Nucleotide Sequences) (Bartlett, S.E. and Davidson, W.S., 1992, BioTechniques, 12(3): 408-411), Total Genomic Probing, MLPs and SLPs.

Despite the wide range of tests available, most are developed for a specific purpose, e.g. to test for the presence of a particular DNA sequence, or to identify individuals e.g. for criminal, paternity or immigration cases or for the protection of rare birds, and are thus limited in their application. Genetic fingerprinting utilises highly polymorphic satellite repeats and whilst excellent for the identification of individuals, specific micro- or mini-satellite repeats are frequently too polymorphic to be used in

speciation studies. RAPD amplifies regions of target DNA randomly but reproducibly using short arbitrary sequence oligonucleotides. Species-specific identification systems using a variety of techniques have been developed (Hu, J. and Quiros, C.F., 1991, Plant Cell Reports, 10: 505-511; Yang, X. and Quiros, C., 1993, Theoretical and Applied Genetics, 86: 205-212; Böhm, M. *et al.*, 1993, Bot. Acta., 106: 448-453) and are widely used (Blackett, R.S. and Keim, P., 1992, Journal of Forensic Sciences, FJCSA, 37(2): 590-596; Forrest, A.R.R. and Carniege, P.R., 1994, Biotechniques, 17: 24-26; Nelson, G. *et al.*, 1992, Textile Research Journal, 62(10): 590-595; Hamlyn, P.F. *et al.*, 1992, J. Text. Inst., 83: 97-103)

However, none is capable of providing a plant-specific characterisation of material such that the phylogeny, in particular the species, of subsequent samples may be identified.

The present inventors have now identified and characterised a polymorphic plant-specific locus having very highly conserved target sequences suitable for amplification (e.g. PCR) primers, the locus showing significant polymorphism at the inter-species level (even among closely related species) but little polymorphic variation at the intra-species level. This has allowed the development of a test to characterise the locus in a wide range of plant species, allowing the subsequent phylogenisation, in particular the speciation, of plant materials.

According to the present invention there is provided a method for determining the phylogeny of plant genetic material comprising at least one intron from at least one RuBisCo (Ribulose 1,5 bisphosphate) small subunit encoding gene,

comprising characterising the intron or introns of the RuBisCo small subunit encoding gene or genes and comparing the said intron or introns to at least one previously characterised intron from a RuBisCo small subunit encoding gene.

Surprisingly, the introns of the RuBisCo small subunit encoding gene (found only in plants) (Dean, C. *et al.*,1989, Ann. Rev. Plant Physiol. Plant Mol. Biol., 40: 415-439) have been found to display (see 'Experimental' section below) significant polymorphism at the inter-species level (even among closely related species) but little polymorphic variation at the intra-species level. This distinction between intra- and inter-species polymorphism appears to be quite stable across a wide range of plant species and this presents a unique method for the phylogenisation, in particular the speciation, of plant genetic material. Intron-based identification or speciation has not previously been performed with the RuBisCo small subunit encoding gene or, in fact, with any other gene.

The characterised intron or introns may be compared to at least one previously characterised intron having substantially the same flanking sequences.

The intron or introns may be sequenced and compared to at least one previously sequenced intron from a RuBisCo small subunit encoding gene.

The method may comprise measuring the sum of the lengths of at least two introns of a RuBisCo small subunit encoding gene or genes and comparing the sum or sums of the lengths of the said introns to the sum of the lengths of at least two previously measured introns. It may comprise measuring the sum of the lengths of at least two

introns of a RuBisCo small subunit encoding gene or genes, together with intervening sequences (i.e. exons). It may comprise measuring the sum of the lengths of all of the introns of a RuBisCo small subunit encoding gene or genes, together with intervening sequences. The method may comprise measuring the sum of the lengths of the introns of the RuBisCo small subunit encoding gene, together with intervening and terminal sequences.

By measuring the sum of the lengths of the introns, in particular by measuring the sum of the lengths of the introns and the intervening sequences, together with any terminal sequences (for example the sequences used to bind amplification primers), a simple set of data may be produced which readily identifies the number of copies of the RuBisCo small subunit encoding gene in the genetic material and also the total lengths of the introns therein.

The length variation of the introns so far observed has been of approximately 1 kb. The number of copies of the gene also varies between species, there so far being found a minimum of approximately 5 copies (tomato) and a maximum of approximately 14 copies (wheat). A simple method of analysis of the size of the introns, whereby the total size of the introns and their intervening sequences is measured (see 'Experimental' below) presents a massive number of possible combinations of readily identified gene copy and total intron size. Other methods which measure the size of just the introns or of particular introns may be equally valid. In addition, the intron or introns may be sequenced and compared to previously sequenced introns - sequences have also been found to display variation which may be used for phylogenisation.

The genetic material may comprise genomic DNA.

The method may be a method for determining the phylogeny of processed plant genetic material. By "processed plant genetic material" is meant plant material containing genetic material which has been subject to degradative processes, e.g. tobacco leaves which have been processed for the production of cigarettes.

It may be a method for determining at least the genus of plant genetic material. It may be a method for determining at least the species of plant genetic material. It may be a method for determining the variety of plant genetic material.

The method may be used to determine whether or not sample plant genetic material is of a particular phylogeny, i.e. whether or not it is of a particular genus or species.

Results so far have indicated that the genus of a plant may be detectable. The results have also shown the clearly-discernable differences between individual species of plants, and in species where there is a particularly high rate of intra-species variation, the variety of plant may also be discernable.

The method may comprise measuring the length of the intron or introns and comparing the length of the said intron or introns to that of at least one previously measured intron.

The plant genetic material may comprise double stranded nucleic acids, the lengths of the introns being measured by:

i) separating the two strands of nucleic acids to produce single-stranded molecules; and

repeating at least once the steps:

ii) annealing to the single-stranded molecules (the "template strands") oligonucleotide primers specific to conserved sequences adjacent to the RuBisCo small subunit encoding sequence or sequences whose length is to be measured;

iii) extending the bound oligonucleotide primers to make complementary copies of the template strands; and

iv) separating the resultant double-stranded products to produce single-stranded molecules,

the phylogeny of the plant genetic material being determined by measuring the lengths of the extended oligonucleotide primers and comparing the lengths of said extended oligonucleotide primers to the length of extended oligonucleotide primers measured from at least one RuBisCo small subunit encoding gene intron or introns.

Two oligonucleotide primers may be used, one specific to a sequence 5' to the first intron of the RuBisCo small subunit encoding gene, the other being specific to a sequence 3' to the last intron of the RuBisCo small subunit encoding gene. The sequence to which the primers are specific is highly conserved.

Hence the invention may be readily worked by measuring the length of the RuBisCo small subunit encoding gene between the 5' junction of the first intron and the 3' junction of the last intron (i.e. between the last intron and exon). This may be done by a simple PCR amplification technique employing a single pair of amplification primers.

Also provided according to the present invention are oligonucleotide primers for use in the method of the present invention. They may be specific to the conserved sequences 5' and 3' of the first and last introns respectively of the RuBisCo small subunit encoding gene. They may have the sequences of SEQ ID NO:1 (5' KSiRGiRTiM RiTGCATGCA GGT) (prime RGI) and/or SEQ ID NO:2 (5' AACAWiGGiA RCTTCCACAT iRYCCAGTA) (primer RGIII) respectively.

Due to the extremely conserved nature of parts of the RuBisCo small subunit encoding gene, such primers may be readily generated (e.g. primers RGI and RGIII, below). They may be designed to allow the amplification in particular of the RuBisCo small subunit encoding gene of, for example, a chosen phylum, class, order or family of plant. Alternatively, they may be designed to allow the amplification in particular of the RuBisCo small subunit encoding gene of a chosen genus of plant.

Also provided according to the present inventionis at least one intron from a RuBisCo small subunit encoding gene adapted for use in a method for determining the phylogeny of plant genetic material, together with the RuBisCo small subunit encoding gene adapted for use in a method of determining the phylogeny of plant genetic material, together with the use of same in said method. Such a method of determining the phylogeny of plant genetic material may be a method according to the present invention.

The invention will be further apparent from the following description, with reference to the several figures of the accompanying drawings, which show, by way of example only, one form of phylogenisation of plant genetic material.

Of the figures:

Figure 1 shows a schematic illustration of a RuBisCo small subunit encoding gene and PCR amplification products. The RuBisCo small subunit encoding gene is shown with two introns (white boxes). Most species have several copies of the gene which contain a variable number of introns that can differ in both length and base composition. By using the conserved coding exons (black boxes) as targets for PCR primers it is possible to amplify a DNA fragment that spans the polymorphic introns of the gene. Several amplification products are produced, one for each copy of the gene. When these fragments are separated by gel electrophoresis a unique species-specific pattern is produced;

Figure 2 shows a schematic illustration of the RuBisCo small subunit encoding gene and primer sequence alignment. The RuBisCo small subunit encoding gene is represented showing the polymorphic introns (white) and the more conserved coding exons (black). Highly conserved regions of the gene were chosen for the design of the PCR primers RGI and RGIII (SEQ ID NOs: 1 and 2 respectively);

Figure 3 shows the genetic profile of thirteen species of *Nicotiana* using the RuBisCo small subunit encoding intron profiling technique. Lanes 1 & 16, size

marker (Pharmacia 100 bp ladder); lane 2, *N. paniculata*; lane 3, *N. glauca*; lane 4, *N. langsdorfii*; lane 5, *N. sanderae*; lane 6, *N. sylvestris*; lane 7, *N. tabacum*; lane 8, *N. repanda*; lane 9, *N. accuminata*; lane 10, *N. nudicaulis*; lane 11, *N. suaveolens*; lane 12, *N. trigonophyla*; lane 13, *N. rustica*; lane 14, *N. solanifolia*; lane 15, No template control;

Figure 4 shows the genetic profile for ten varieties of *Nicotiana rustica*. Lanes 1 & 13, size marker (Pharmacia 100 bp ladder); lane 2, *N. rustica - var.* Souffi Sétif; lane 3, *N. rustica - var.* Souffi Gabès; lane 4, *N. rustica - var.* Pumilag; lane 5, *N. rustica - var.* Limonka; lane 6, *N. rustica - var.* Kapa; lane 7, *N. rustica - var.* Souffi Cap Bon; lane 8, *N. rustica - var.* Ovata; lane 9, *N. rustica - var.* Texana F; lane 10, *N. rustica - var.* Brasilia; lane 11, *N. rustica - var.* Asiatica; lane 12, No template control;

Figure 5 shows the genetic profile for thirteen species from different taxonomic groups. Lanes 1 & 16, size marker (Pharmacia 100 bp ladder); lane 2, *Eucaliptus urophylla*; lane 3, *Eucaliptus nitens*; lane 4, Walnut (*Juglans spp.*); lane 5, Turkey Oak (*Quercus cerris*); lane 6, Tomato (*Lycopersicum spp.*); lane 7, Potato (*Solanum tuberosum*); lane 8, Pea (*Pisum sativum*); lane 9, Rice (*Oryza satvia*); lane 10, Cacao (*Theobroma cacao*); lane 11, *Arabadopsis spp.*; lane 12, Soybean; lane 13, Maize (*Zea mays*); lane 14, Swede; lane 15, No template control; and

Figure 6 shows the genetic profile analysis of sonicated genomic DNA. Gel (A) shows the simulated degradation of *Nicotiana tabacum* DNA using sonication. Lane 1, size marker, λ *Hin* dIII digest (23130, 9416, 6557, 4361, 2322, 2027, 564, 125 bp); Lane 2, untreated genomic DNA; lanes 3, 30 seconds; lane 4, 60 seconds; lane 5, 90 seconds; lane 6, 120 seconds; lane 7, 150 seconds; lane 8, 180 seconds. Gel (B) shows

PCR of sheared DNA from Gel (A) using standard reaction conditions. Lane 1, size marker, 100 bp ladder (Pharmacia); Lanes 2 to 8, amplification of sheared DNA of corresponding lane in Gel (B) ;lane 9, no template control.

Experimental

Materials and Methods

Plant Materials

The following species of *Nicotiana* were obtained from Le Institute Du Tabac Bergerac, France. *N. rustica - var.* Asitica ; *N. rustica - var.* Brasilia ; *N. rustica - var.* Kapa ; *N. rustica - var.* Limonka ; *N. rustica - var.* Ovata ; *N. rustica - var.* Pumilag ; *N. rustica - var.* Souffi Cap Bon ; *N. rustica - var.* Souffi Gabès ; *N. rustica - var.* Souffi Sétif ; *N. rustica - var.* Texana F; *N. paniculata* L ; *N. knightiana* Goodsp.; *N. raimondii* Macbr ; *N. cordifolia* Phil ; *N. solanifolia* Walp ; *N. glauca* Grah.; *N. tabacum* L.; *N. trigonophylla* Dun.; *N. repanda* Willd. ; *N. acuminata* Grah ; *N. nudicaulis* Wats. ; *N. suaveolens* Lehm. Several species of Eucalyptus were kindly donated by Advanced Technologies (Cambridge) (Cambridge Science Park, England). Other samples of fresh and processed tissues were purchased from a local supermarket.

Extraction of genomic DNA

No single DNA extraction procedure provided pure DNA from all tissues examined. The technique developed by Doyle & Doyle (Doyle, J.J. and Doyle J.L., 1991, Biotechniques (focus), 12(1): 13-15) was found to yield high concentrations of PCR amplifiable DNA from most plant sources. However, for difficult tissues containing high concentrations of phenolic compounds and/or starch this extraction protocol was combined with a further DNA purification step using the silica based protocol developed by Höss (Höss, M. and Paabo, S., 1993, Nucleic Acid Res., 21: 3913-3914).

Sequence alignments

RuBisCo small subunit encoding gene sequences were extracted from the EMBL database and analysed using Screen Orientated Multiple Alignment Procedure (SOMAP) software (Parry-Smith, D. J. and Attwood, T.K., 1991, CABIOS, 7(2): 233-235). A sequence alignment of *Nicotiana sylvestris*, *Nicotiana plumbaginifolia*, *Nicotiana tabacum*; pine (*Pinus*); pea (*Pisum sativum*); maize (*Zea mays*); clover (*Trifolium spp.*); pear (*Pyrus communis*) and larch (*Larix spp.*) was used to locate conserved sequence regions and design PCR primers (Figure 2).

Oligonucleotide synthesis

Oligonucleotides were synthesised on an Applied Biosystems 392 DNA synthesis instrument (Perkin Elmer). Following cleavage from the solid support and deprotection, no further purification was necessary and oligonucleotides were stored in 80% ammonia at -20°C.

PCR conditions

Reactions were performed in PCR buffer (10 mM Tris pH 8.0, 0.01% gelatin, 0.1% Triton X 100, 80 mM KCl; 3.5 mM MgCl) containing 35 pM of each primer, 250 µM each dNTP (Pharmacia Ultra pure), 20 to 100 ng target DNA, 5 Units *Taq* (Boehringer Mannheim). Reaction volumes were made up to 25 µl (HPLC grade water (Sigma)) and overlayed with 1 drop of mineral oil (Sigma). Amplification of the RuBisCo small subunit encoding genes was achieved by an initial denaturation step at 94°C, 3 minutes followed by addition of the polymerase and 35 cycles of 94°C, 20 sec; 63.5°C, 20 sec; 72°C, 60 sec with a final extension step of 3 minutes at 72°C using a Hybaid Omni gene Thermal Cycler with simulated tube control (calibration factor 150).

PCR product analysis

Aliquots (3 to 10 µl) of the PCR products were fractionated through 2% agarose 1:1 Sigma & Nusieve (RTM) (3:1 agarose; FMC) for 2 hours at 7 v/cm in Tris-Borate (TBE) buffer and stained with ethidium bromide as described by Sambrook (Maniatis, T., Fritsch, E.F. and Sambrook, J., 1982, Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, Cold Spring Harbor). DNA gels were photographed using an Olympus OM4 Ti with Ilford XP2 400 film. DNA fragment sizes were determined by comparison to λ *Hin* dIII digest and Pharmacia 100 bp ladder size markers. DNA fragments were isolated from the PCR reaction by direct ligation using PCR, TA cloning kit (Invitrogen). DNA from recombinant plasmids was prepared for sequencing using the Wizard Miniprep Kit (Promega). Sequencing reactions were conducted using the purified plasmid as template at Kings College School of Medicine on an automated ABI 373A sequencer (Perkin Elmer).

Hybridisation

PCR products were randomly labelled using [α^{32}P] dCTP and the mega prime kit (Amersham). Southern analysis was carried out as described by Sambrook (Maniatis, *supra*). Oligonucleotide probes were end-labelled using [χ^{32}P] ATP and a Boehringer Mannheim end-labelling kit. Radiolabelled probes were separated from unincorporated nucleotides using Pharmacia G50 columns.

Simulated Degradation of Genomic DNA

Freshly extracted *Nicotiana tabacum* genomic DNA was exposed to high frequency sonic waves using a Heat Systems ultrasonic processor coupled to a 20 KHz ultrasonic convertor. The ultrasonic probe was immersed in ice water and the tube containing the

DNA held adjacent to the probe in the path of the sonic output. The tube was briefly vortexed and centrifuged before removal of any DNA.

Results & Discussion

Amplification from thirteen species of *Nicotiana* (Figure 3) demonstrates that the Ribulose-1,5-Bisphosphate Carboxylase genes (i.e. the various copies of the RuBisCo small subunit encoding gene in a plant) contain sufficient sequence variation to distinguish between closely related plant species. By comparing the band pattern, including band number, sizes and relative intensities, with the pattern produced from a reference sample it is possible to identify unknown *Nicotiana* samples. Gene structure and sequence data (Dean, C. *et al.*,1989, Ann. Rev. Plant Physiol. Plant Mol. Biol., 40:415-439) allows prediction that PCR amplification would produce two distinct subgroups of amplification products. Larger amplification products with a size in excess of 700 bp, represent gene fragments containing three introns, while smaller amplification products, with a size of less than 600 bp, probably contain only two introns.

To investigate the extent of varietal differences the Ribulose-1,5-Bisphosphate Carboxylase gene profile of ten varieties of *Nicotiana rustica* were analysed (Figure 4). Some polymorphism exists, with *Nicotiana rustica - var.* Texana, *var.* Brazilia and *var.* Asiatica apparently possessing an additional copy of the gene which produces an amplification product with a molecular mass of 880 bp. Variation at the test loci within *Nicotiana* species barriers is evidence that Ribulose-1,5-Bisphosphate Carboxylase intron amplification could be used to provide a form of identification beyond the species level to the variety level. However, the level of polymorphism in any species of interest should

be determined in advance of any routine testing. If, for example, a test were to be conducted upon an unknown variety of *Nicotiana rustica* and the result did not match that of any previously determined samples this could mean that the unknown tissue was a member of a variety which had not yet been analysed but was still *Nicotiana rustica*. The level of polymorphism is likely to vary between different species and the results presented provide only an indication of the level of polymorphism which may occur.

The amplification products for the *Nicotiana rustica* varieties (Figure 4) were digested individually with the four base cutting restriction enzymes *Rsa*I and *Alu*I. The bands generated for the ten varieties were very similar, allowing the identification of particular varieties and groups of varieties (results not shown) and demonstrating that there is little sequence variation within the amplification products. Hybridisation studies to investigate the use of the amplification products as species-specific probes also showed that there is a high degree of sequence homology between closely related species. The largest amplification product of *Nicotiana glauca* (605 bp) was labelled and hybridised to amplification products form other *Nicotiana* species. Whilst the probe did not hybridise to bands for most other species, even under stringent wash conditions it did hybridise to *Nicotiana langsdorfii* and *Nicotiana sanderae*, two species which appear to be closely related to *Nicotiana glauca*. The evidence of sequence conservation between closely related species suggests that individuals of a given species will contain a low level of polymorphism.

The PCR conditions have been applied to a wide range of different species. As shown in Figure 5, all samples yield banding patterns and they show similarities for related species and characteristic variation between different plants. Again the amplification

pattern produced was repeatable and could be used to identify the original DNA sample. Several of the amplification products have been sequenced and the high levels of homology show them to be the Ribulose-1,5-Bisphosphate Carboxylase genes.

DNA extraction from commercially processed plant tissues has shown that any residual nucleic acid is often highly degraded. In an effort to simulate commercial processing in a controlled manner high molecular weight DNA, isolated from fresh tissue, was sonicated for varying lengths of time to produce a range of DNA samples with decreasing average fragment size (Figure 6A). When tested with the standard reaction conditions an identical amplification profile is produced as for fresh tissue (Figure 6B), suggesting that this type of analysis may be applicable to the identification of commercially processed plant material or other samples from which only poor quality DNA is available. To test this hypothesis DNA was isolated from several brands of cigarettes and cigars and tested under the standard reaction conditions. All samples gave the genetic profile of *Nicotiana tabacum* (results not shown) despite the condition of the DNA, which was highly degraded.

DISCUSSION

The reaction conditions outlined in this study are capable of reproducibly amplifying a polymorphic section of the RuBisCo small subunit encoding gene from over twenty different species. It has been demonstrated that the level of polymorphism which occurs naturally at these loci are generally sufficient to allow differentiation between any two different species following gel separation of the PCR amplification products. While some genetic variation exists in the introns of the Ribulose-1,5-Bisphosphate Carboxylase

genes within a species, as shown by the low but detectable polymorphisms between *Nicotiana rustica* varieties, this could be beneficial to any studies where the level of polymorphism had already been determined. As the target loci are present only in the plant genome the problems of spurious amplification due to human contamination of the sample is significantly decreased, whilst still allowing the detection of contaminating plant matter in adulterated samples. Specific loci amplification of this nature therefore partially avoids one of the fundamental drawbacks associated with PCR based analytical techniques. Kocher (Kocher, T. D. *et al.*, 1989, Proc. Natl. Acad. Sci. U.S.A., 86: 6196-6200) developed a set of primers suitable for amplifying a section of the cytochrome B gene. Many studies have utilised these reaction conditions to amplify the gene from many members of the animal kingdom, and have used the information gained from subsequent analysis of the amplification product in phylogenetic studies.

Amplification of the RuBisCo small subunit encoding gene, in particular the regions comprising at least its introns, provides a useful additional tool for analysis in the plant kingdom.

CLAIMS

1. A method for determining the phylogeny of plant genetic material comprising at least one intron from at least one RuBisCo small subunit encoding gene, comprising characterising the intron or introns of the RuBisCo small subunit encoding gene or genes and comparing the said intron or introns to at least one previously characterised intron from a RuBisCo small subunit encoding gene.

2. A method according to claim 1 wherein the characterised intron or introns are compared to at least one previously characterised intron having substantially the same flanking sequences.

3. A method according to either one of claims 1 or 2 wherein the intron or introns are sequenced and compared to at least one previously sequenced intron from a RuBisCo small subunit encoding gene.

4. A method according to any one of the preceding claims wherein it comprises measuring the length of the intron or introns and comparing the length of the said intron or introns to that of at least one previously measured intron.

5. A method according to either one of claims 3 or 4 wherein it comprises measuring the sum of the lengths of at least two introns of a RuBisCo small subunit encoding gene or genes and comparing the sum or sums of the lengths of the said introns to the sum of the lengths of at least two previously measured introns.

6. A method according to claim 5 wherein it comprises measuring the sum of the lengths of at least two introns of a RuBisCo small subunit encoding gene or genes, together with intervening sequences.

7. A method according to claim 6 wherein it comprises measuring the sum of the lengths of all of the introns of a RuBisCo small subunit encoding gene or genes, together with intervening sequences.

8. A method according to any one of the preceding claims wherein the genetic material comprises genomic DNA.

9. A method according to any one of the preceding claims wherein it is used to determine the phylogeny of processed plant genetic material.

10. A method according to any one of the preceding claims wherein it is a method of determining at least the genus of plant genetic material.

11. A method according to claim 10 wherein it is a method of determining at least the species of plant genetic material.

12. A method according to claim 11 wherein it is a method of determining the variety of plant genetic material.

13. A method according to any one of the previous claims, the plant genetic material comprising double stranded nucleic acids wherein the lengths of the introns are measured by:

i) separating the two strands of nucleic acids to produce single-stranded molecules; and

repeating at least once the steps:

ii) annealing to the single-stranded molecules (the "template strands") oligonucleotide primers specific to conserved sequences adjacent to a RuBisCo small subunit encoding gene intron or introns whose length is to be measured;

iii) extending the bound oligonucleotide primers to make complementary copies of the template strands; and

iv) separating the resultant double-stranded products to produce single-stranded molecules,

the phylogeny of the plant genetic material being determined by measuring the lengths of the extended oligonucleotide primers and comparing the lengths of said extended oligonucleotide primers to the length of extended oligonucleotide primers measured from at least one RuBisCo small subunit encoding gene intron or introns.

14. A method according to claim 13 wherein two oligonucleotide primers are used, one specific to a conserved sequence 5' to the first intron of a RuBisCo small subunit encoding gene, the other being specific to a conserved sequence 3' to the last intron of a RuBisCo small subunit encoding gene.

15. Oligonucleotide primers for use in the method of either one of claims 13 or 14.

16. Oligonucleotide primers according to claim 15 wherein they have the sequences of at least one of the group of SEQ ID NO:1 and SEQ ID NO:2.

17. At least one intron from a RuBisCo small subunit encoding gene adapted for use in a method for determining the phylogeny of plant genetic material.

18. The RuBisCo small subunit encoding gene adapted for use in a method for determining the phylogeny of plant genetic material.

19. At least one intron from a RuBisCo small subunit encoding gene according to either one of claims 17 or 18 adapted for use in a method for determining the phylogeny of plant genetic material according to any one of claims 1-14.

20. The use of at least one intron from a RuBisCo small subunit encoding gene according to either one of claims 17 or 18 in a method of determining the phylogeny of plant geneic material.

The Patent Office

22

Application No: GB 9703774.1 **Examiner:** Colin Sherrington
Claims searched: 1 to 14, 16, 20 **Date of search:** 7 May 1997

Patents Act 1977
Search Report under Section 17

Databases searched:

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.O): C3H(HB4A); G1B(BAC)

Int Cl (Ed.6): C12Q 1/68

Other: ONLINE: WPI,CLAIMS,DIALOG/BIOTECH

Documents considered to be relevant:

Category	Identity of document and relevant passage	Relevant to claims
A	C.R.Acad.Sci.Paris,Sciences de la vie/Life sciences 1994,317,685-692 - Ludovic Gielly *et al.* "Chloroplast DNA polymorphism at the intrageneric level and plant phylogenies"	1
A	Mol.Biol.Evol. 1994,11(5),769-777 -Ludovic Gielly *et al.* "The Use of Chloroplast DNA to Resolve Plant Phylogenies: Noncoding versus *rbcL* Sequences"	1
A	Mol.Gen.Genet. 1993,239,450-462 -E.Jay DeRocher *et al.* "The six genes of the Rubisco small subunit multigene family from *Mesembryanthemum crystallinum*, a facultative CAM plant"	1
A	Annu.Rev.Plant Physiol.Plant Mol.Biol. 1989,40,415-439 -Caroline Dean *et al.* "Structure, Evolution, and Regulation of *RbcS* Genes in Higher Plants"	1
A	Aust.J.Bot. 1983,31,395-409 -P.G.Martin *et al.* "The Study of Plant Phylogeny using Amino Acid Sequences of Ribulose-1,5-Bisphosphate Carboxylase. I Biochemical Methods and the Patterns of Variability"	1

X Document indicating lack of novelty or inventive step
Y Document indicating lack of inventive step if combined with one or more other documents of same category.
& Member of the same patent family
A Document indicating technological background and/or state of the art.
P Document published on or after the declared priority date but before the filing date of this invention.
E Patent document published on or after, but with priority date earlier than, the filing date of this application.